

Tables of Contents

1.0) Abstract	2
2.0) Introduction	2
3.0) Background	3
3.1) Data Cycle and Datafication	3
3.2) K-Means	3
4.0) Research Question and Hypothesis	4
5.0) Research Objectives	5
6.0) Research Significances	6
7.0) Dataset	7
8.0) Literature Review	
9.0) Analysis	
10.0) Case Study and K-Means	
11.0) Comparison of Covid-19 case between Philippines and Malaysia	
12.0) References	

1.0) Abstract

Utilizing big data analytics and machine learning techniques can be an excellent choice in predicting the pattern of Covid-19 cases in the Philippines. The information obtained would later help the frontlines, the government, and citizens to prepare for the worst-case scenario possible in the later future. We are going to predict whether the chances of an individual being infected by Covid-19 have to do with the individual's age, sex, or province. Analyzing these data would then help guide further precautions towards these highly vulnerable individuals. The machine learning algorithm that we would utilize is the K-Means Clustering Algorithm. We choose this clustering algorithm because we would like to group our datasets into groups that have a high risk being infected by Covid 19 and those who do not. Clustering them into groups will help us rule out the ones that could be victims and possibly lead them to death.

2.0) Introduction

As we all know, the universe has been hit with a pandemic called the Covid-19 virus. Countries all over the world were highly affected by Covid-19. As months went by, some countries were able to flatten the curve and some were not able to and still worked hard to fight the virus. We are interested in studying Covid-19 in the Philippines as there was a recent case regarding some Filipinos that came to Malaysia and were tested positive for Covid19. Plus, there were huge differences in statistical reports about a few months ago in the Philippines, and this country records the region's biggest daily rise in Covid-19 deaths.

As stated from worldometer info website, there are 1571317 cases from Southeast Asia and the Philippines is ranked as the top Covid-19 cases in this region. The Philippines recorded 2200 new infections and 28 deaths which makes a total of 74390 cases consisting of 24383 recovered, 1871 deaths and 351 in the Incentive Care Unit (ICU).

Utilizing big data and machine learning techniques can be an excellent option in predicting which class will get infected by Covid19 viruses easily and which class can recover faster. By examining the victims' age and current conditions we can predict their conditions if they are infected with the Covid-19 virus. In this study, we do not include other effects that may

cause the infected victim's conditions to get worse such as having a history with serious diseases, and others. We will train the data and test it with our model. This model would later help the frontlines, to lessen the numbers of death and to decrease the number of new cases daily.

3.0) Background

3.1) Data Cycle and Datafication

Using the Covid-19 in the Philippine dataset provided by this website <https://www.kaggle.com/cvronao/covid19-philippine-dataset> we can provide a prediction model that can predict some statements based on our researcher.

3.2) K-Means

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. A cluster refers to a collection of data points aggregated together because of certain similarities. We will define a target number k , which refers to the number of centroids that are needed in the dataset. A centroid is the imaginary or real location representing the center of the cluster. K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible.

This algorithm is faster in computationally than hierarchical clustering if variables are huge and keep the k value small. K-means is simple to implement. Plus, K-Means produces tighter clusters than hierarchical clustering, especially if the clusters are globular.

However, it is difficult to predict K-value since we need to predict it manually. K-Means are also unable to handle noisy data and outliers.

4.0) Research Question and Hypothesis

Research question(s) are the most important part of the research. It acts as a catalyst for the researchers to stay focused on the study, resolve on a methodology that suits the research and also gives a base view of what the research would cater to. There are three kinds of research questions, which are descriptive - the essence of describing the study, comparative - analyzing

the effects between two or more variables on the dependent variable and casual - whether the dependent variable could cause other outcomes. Therefore, based on these three kinds of questions, we have proposed three questions with respect to our research:-

- I. Philippine has recorded Southeast Asia's biggest daily jump in coronavirus deaths in the month of June. Why?
- II. Which age group is the most affected with Covid-19?
- III. Which age group has the highest risk of death when affected with covid-19?
- IV. Covid-19 in the Philippine has reduced in April but increased again in May. Which province in the Philippine has led to the increased number of cases in the Philippine?

5.0) Research Objectives

- I. To determine the cause that led to an increased number of cases in the month of May in the Philippines.
- II. To determine the main factor that caused the biggest daily jump of deaths due to Covid-19 in the month of June.
- III. To evaluate which age group are most likely to be infected by Covid-19.
- IV. To understand whether the chances of a person being infected by Covid-19 has anything to do with their age, sex, or province.

6.0) Research Significances

As included in our objectives for this COVID-19 in Philippines research, we want to find out information related to spike in cases, correlations and reasons for them, age groups with highest risk, whether location has effect on infection chances and looking at the increase in a particular province. Finding these out will be very useful and very significant for a lot of reasons.

The most important being, if we find a cause of spike in cases, policy makers can make policy and curb the increase and take preventative measures for the future. Public can take safety precautions understanding what it is that made the cases to increase rapidly. They can try and avoid doing things that put them at risk.

Finding out the age group most likely to get COVID-19 will help to narrow down and taking measures against more infections by focusing more on them. Looking at everyone equally will mean people who really need it, will not be taken into matters urgently. Not everyone is equal at risk and so resources should be allocated according to their need.

People in charge can understand and take actions like lockdown, semi-lockdown, controlling movement and such, more effectively if they know whether a location like provinces within the Philippines has a lot to do with infections. Because if it does, then they can take action according to each province and each place, like placing lockdowns on provinces that are hardly hit and letting people have certain freedoms if they are in stable provinces. As we can see, the research can show a lot of useful information that can be used for the good of people against the pandemic that is COVID-19.

7.0) Dataset

<https://www.kaggle.com/cvronao/covid19-philippine-dataset>

We obtained our dataset from the links above. The title of the dataset is COVID-19 Cases in the Philippines. It contains 12092 rows including the column attribute names rows and 16 columns. However, there are two columns which are health_status and status that have the same data. Therefore, we have decided to drop the health_status column during data cleaning. This collection of datasets made available to the public by the Philippine Department of Health (DOH) through their Google Sheets data drop. Below is the screenshot of part of the dataset not including health_status column:

This data set provides the age of the patient and also the age group. There are two kinds of gender, which is categorised as females and males. There are five columns that have a date format which is date_announced where it is the official date for a patient that was diagnosed as positive COVID, date_recovered where it is a date the patient recover from COVID, date_of_death where it is a date when a patient dies due to the COVID, date_announced_as_removed where it is a date for closing COVID case of a patient, and

date_of_onset_of_symptom is the date of the day when the patient had the COVID symptom. Province, region, and muni_city columns provide the geolocation or the area of the patient coming from. The status column is categorized as recovered and died while the home_quarantine column is categorised as Yes for those who are staying at home for quarantine or No for those who are not staying at home.

Case_id	age	age_group	sex	date_announced	date_recovered	date_of_death	status	date_announced_as_removed	province	muni_city	Home_quarantine	date_of_onset_of_symptoms	pregnant	region
C404174	38	35 to 39	Female	30/1/2020	NA	NA	Recovered	7/2/2020	Negros Oriental	Dumaguete City	NA	21/1/2020	NA	Central Visayas (Region VII)
C462688	44	40 to 44	Male	3/2/2020	NA	1/2/2020	Died	2/2/2020	Negros Oriental	Dumaguete City	NA	18/1/2020	NA	Central Visayas (Region VII)
C387710	60	60 to 64	Female	5/2/2020	31/1/2020	NA	Recovered	5/2/2020	Bohol	Panglao	No	21/1/2020	NA	Central Visayas (Region VII)
C377460	48	45 to 49	Male	6/3/2020	NA	NA	Recovered	27/3/2020	Metropolitan Manila	Taguig	No	3/3/2020	NA	Metropolitan Manila
C498051	62	60 to 64	Male	6/3/2020	NA	11/3/2020	Died	12/3/2020	Rizal	Cainta	No	25/2/2020	NA	CALABARZON (Region IV-A)
C130591	58	55 to 59	Female	7/3/2020	NA	12/3/2020	Died	12/3/2020	Rizal	Cainta	No	27/2/2020	NA	CALABARZON (Region IV-A)
C178743	39	35 to 39	Male	8/3/2020	21/3/2020	NA	Recovered	22/3/2020	Metropolitan Manila	Makati City	No	3/3/2020	NA	Metropolitan Manila
C202135	58	55 to 59	Male	8/3/2020	NA	NA	Recovered	25/3/2020	Metropolitan Manila	Quezon City	NA	20/2/2020	NA	Metropolitan Manila
C440075	33	30 to 34	Male	8/3/2020	5/4/2020	NA	Recovered	9/4/2020	Metropolitan Manila	Pasig City	Yes	1/3/2020	NA	Metropolitan Manila
C557002	86	80+	Male	8/3/2020	NA	14/3/2020	Died	15/3/2020	Metropolitan Manila	Marikina	No	1/3/2020	NA	Metropolitan Manila

8.0) Literature Review

From tropical medicine and health websites reported that the first case in the Philippines was by a Chinese nationals couple that was on vacation in the Philippines during January 2020. The first patient is a 39-year-old female and was admitted to San Lazaro Hospital in Manila on the 25th January 2020 while the second patient is a 44-year-old male.

The presence of Covid-19 in low- and middle-income countries such as the Philippines is raising important concerns about effective pandemic response and preparedness in terms of health systems and the pervasiveness of misinformation. Certain places in the Philippines have poor resource settings that might not have access to regular and reliable sources of information about disease. The people who stay around the areas may not notice the symptoms and

information about disease etiology. This also may be a factor why Covid 19 cases in the Philippines increase.

Journal from (Lauer, Grantz, Bi, Jones, 2020) stated that from 10,000 cases of Covid 19 there are 101 cases that will develop symptoms after 14 days of quarantine. The authors also mentioned that the most effective duration to stay quarantined is 14 days which is 99% effectiveness. The WHO also established the quarantine period must be 14 days. This is why Malaysia Ministry of Health implements 14 days of quarantine for those who have the symptoms and for those who just come back from the high case recorded countries. Quarantine is one way to prevent the outbreaks of Covid 19 virus.

9.0) Analysis

9.1) Data Cleaning

Some of the data cleaning processes that we have done throughout the whole process of analysis are changing data types, drop columns, remove NAN and drop duplicates.

1. Change data types for all columns to the correct data type

df.dtypes			
case_id	object	case_id	category
age	float64	age	float64
age_group	object	age_group	category
sex	object	sex	category
date_announced	object	date_announced	datetime64[ns]
date_recovered	object	date_recovered	datetime64[ns]
date_of_death	object	date_of_death	datetime64[ns]
status	object	status	category
date_announced_as_removed	object	date_announced_as_removed	datetime64[ns]
province	object	province	category
muni_city	object	muni_city	category
health_status	object	health_status	category
home_quarantined	object	home_quarantined	category
date_of_onset_of_symptoms	object	date_of_onset_of_symptoms	datetime64[ns]
pregnant	object	pregnant	category
region	object	region	category
dtype: object		dtype: object	

Original data type

Correct data type format

2. Drop health_status column as it has the same data with status column

```
print(df.keys())

Index(['case_id', 'age', 'age_group', 'sex', 'date_announced',
      'date_recovered', 'date_of_death', 'status',
      'date_announced_as_removed', 'province', 'muni_city', 'health_status',
      'home_quarantined', 'date_of_onset_of_symptoms', 'pregnant', 'region'],
      dtype='object')
```

All columns in the dataset

```
print(df.keys())

Index(['case_id', 'age', 'age_group', 'sex', 'date_announced',
      'date_recovered', 'date_of_death', 'status',
      'date_announced_as_removed', 'province', 'muni_city',
      'home_quarantined', 'date_of_onset_of_symptoms', 'pregnant', 'region'],
      dtype='object')
```

3. Remove NAN

	date_announced	home_quarantined
0	2020-01-30	NaN
1	2020-02-03	NaN
2	2020-02-05	No
3	2020-03-06	No
4	2020-03-06	No
...
12086	2020-05-15	NaN
12087	2020-05-15	NaN
12088	2020-05-15	NaN
12089	2020-05-15	Yes
12090	2020-05-15	NaN

12091 rows × 2 columns

Total rows for date_announced and home_quarantined columns before remove NAN values is 12091. Total rows for columns date_announced and home_quarantined after remove NAN is 5310.

	date_announced	home_quarantined
0	2020-02-05	No
1	2020-03-06	No
2	2020-03-06	No
3	2020-03-07	No
4	2020-03-08	No
...
5305	2020-05-15	Yes
5306	2020-05-15	Yes
5307	2020-05-15	Yes
5308	2020-05-15	Yes
5309	2020-05-15	Yes

5310 rows × 2 columns

4. Drop duplicates based on age_group

	age	age_group	count
0	38.0	35 to 39	1061.0
1	44.0	40 to 44	967.0
2	60.0	60 to 64	921.0
3	48.0	45 to 49	1002.0
4	62.0	60 to 64	921.0
...
12065	35.0	35 to 39	1061.0
12066	37.0	35 to 39	1061.0
12067	62.0	60 to 64	921.0
12068	18.0	15 to 19	282.0
12069	50.0	50 to 54	973.0

12070 rows × 3 columns

The rows before the delete the duplicates based on age_group

	age	age_group	count
4016	2.0	0 to 4	154.0
10040	10.0	10 to 14	169.0
9927	18.0	15 to 19	282.0
10340	24.0	20 to 24	767.0
3631	28.0	25 to 29	1289.0
4630	30.0	30 to 34	1510.0
10940	35.0	35 to 39	1061.0
9555	42.0	40 to 44	967.0
6820	45.0	45 to 49	1002.0
11309	8.0	5 to 9	122.0
10074	51.0	50 to 54	973.0
3047	57.0	55 to 59	967.0
1352	62.0	60 to 64	921.0
2599	65.0	65 to 69	713.0
2775	74.0	70 to 74	565.0
50	79.0	75 to 79	287.0
11054	87.0	80+	321.0

Rows left after duplicates has been deleted based on age_group

9.2) Question 1: Philippine has recorded Southeast Asia's biggest daily jump in coronavirus deaths in the month of June. Why?

First we started to analyze the data by calculating the percentage of home quarantine as home quarantined play a crucial role in preventing the spreading of Covid-19.

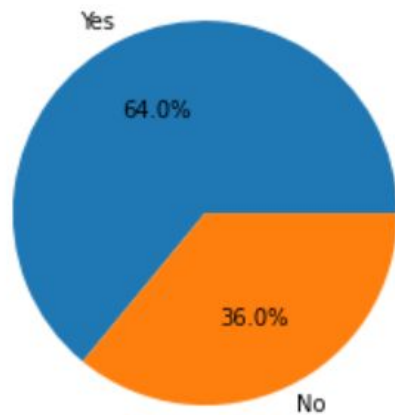
Using formula to calculate the percentage of Yes and No for people quarantine at home

```
total = quarantined_yes + quarantined_no
yes_percentage = round((quarantined_yes / total)*100)
yes_percentage
```

```
date_announced    64.0
home_quarantined    64.0
dtype: float64
```

```
[ ] no_percentage = 100 - yes_percentage
no_percentage
```

```
date_announced    36.0
home_quarantined    36.0
dtype: float64
```



In Philipphine the percentage of people who quarantine from February until May is only 64%.

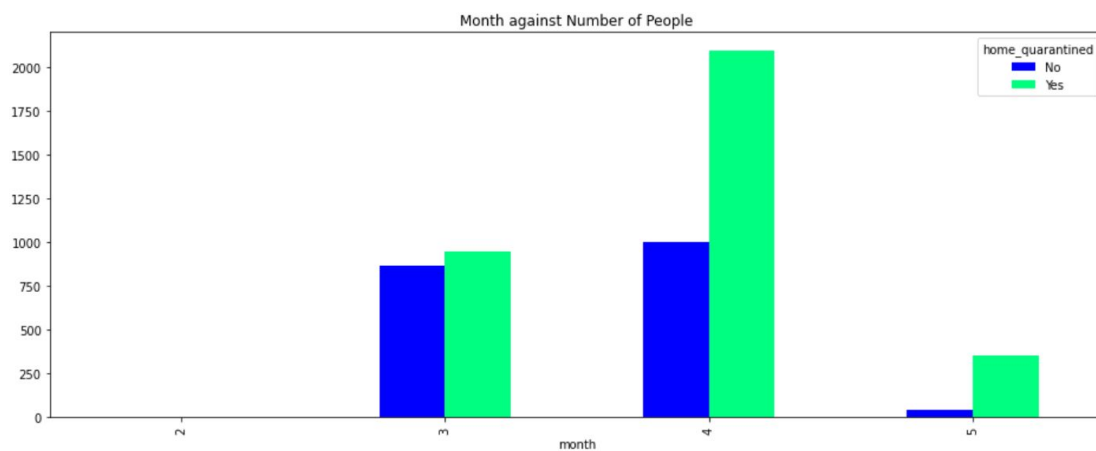
home_quarantined	No	Yes
month		
2	1	0
3	865	946
4	1003	2098
5	41	356

By using the unstack function, we have separated No and Yes data from home_quarantined columns and accumulate the number of Yes and No according to the respective month. We can see that the number of people who do not quarantine from March increases until April as the number of cases increase. Thus, we can assume that quarantined plays a crucial factor that leads to an increase number of Covid-19 case in Philipphine on April and continue to increase until June.

To see the flow of cases according to month we have extracted month from the date_announced column.

	date_announced	home_quarantined	month
0	2020-02-05	No	2
1	2020-03-06	No	3
2	2020-03-06	No	3
3	2020-03-07	No	3
4	2020-03-08	No	3
...
5305	2020-05-15	Yes	5
5306	2020-05-15	Yes	5
5307	2020-05-15	Yes	5
5308	2020-05-15	Yes	5
5309	2020-05-15	Yes	5

5310 rows × 3 columns

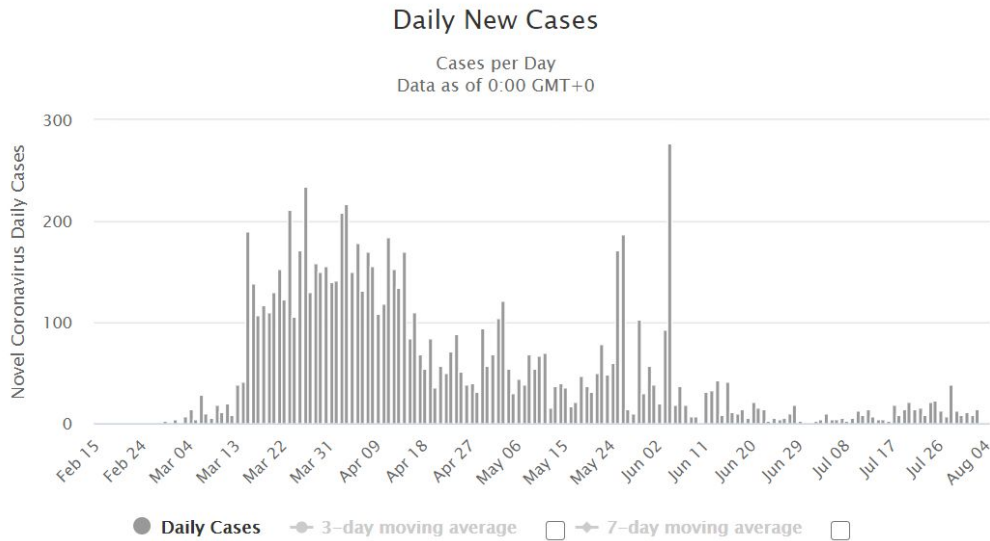


The graph shows there is a huge gap between the number of cases in March and number of cases in April.



This graph was taken from the Philippines Department of Health website <https://www.doh.gov.ph/covid19tracker>. The cases started to increase in April and have rapidly increased in July. From Channel News Asia, (July, 2020) the Philippines recorded its largest daily spike in Covid-19 deaths on Monday (Jul 13). The number of Covid-19 infections has more than tripled to 56,259 since June 1 and continues to rise until the beginning of August when the government began to lighten tight restrictions, including enabling public transportation, restaurants, and malls to open up at restricted capacity to restart the economy.

In Malaysia, After 100 days after the movement control order (MCO) came into effect on March 18 and we can see the country's progress in flattening the Covid-19 infection curve, thereby reducing population transmission of the virus. From The Star, Health Director-General Datuk Dr. Noor Hisham Abdullah said the drastic actions of the government at that time restricting people's movement and closing borders gave the Ministry of Health the opportunity to increase the ability of public health, labs, health clinics, and hospitals, thus flattening the Covid-19 curve. In June, the number of reported cases fell to just 208, and mortality rates fell to 1.4% of the total number of cases. This shows how important home quarantine is in preventing the spreading of the virus.

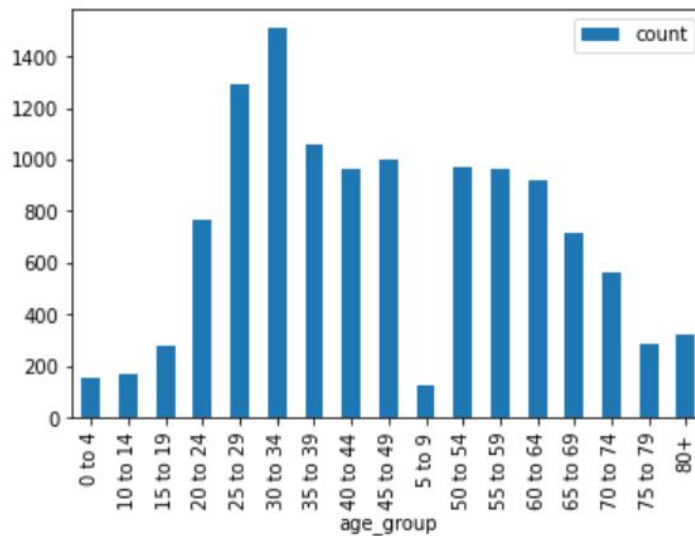


This graph was taken from www.worldometers.info/coronavirus. It shows the effectiveness of home quarantined when the number of daily new cases has declined during the MCO.

9.3) Question 2: Which age group is the most affected with Covid-19?

To see which age group has the most cases for Covid-19, we have grouped all the cases according to the age group.

age_group			count	8	45 to 49	1002.0
0	0 to 4	154.0		9	5 to 9	122.0
1	10 to 14	169.0		10	50 to 54	973.0
2	15 to 19	282.0		11	55 to 59	967.0
3	20 to 24	767.0		12	60 to 64	921.0
4	25 to 29	1289.0		13	65 to 69	713.0
5	30 to 34	1510.0		14	70 to 74	565.0
6	35 to 39	1061.0		15	75 to 79	287.0
7	40 to 44	967.0		16	80+	321.0



From the graph, the age group of 30 to 34 has the highest cases of Covid-19. We have divided the age group to a smaller group age, and we also have combined the count for the total case from the age_group column according to the new age group.

Age	Age Group	Total Case
0 - 4 years old	Infant	154
5 - 14 years old	Children	291
15 - 19 years old	Teenagers	282
20 - 34 years old	Young adult	3566
35 - 54 years old	Middle age	4003
55 years old and older	Older adulthood	3774

Based on the analysis, we can conclude that middle age and older adulthood age groups are the most highly affected by Covid-19 in the Philippines.

9.4) Question 3: Which age group has the highest risk of death when affected with Covid-19?

To find which age group has the highest risk of death when affected with Covid-19 we have selected to do the analysis for age_group and status columns.

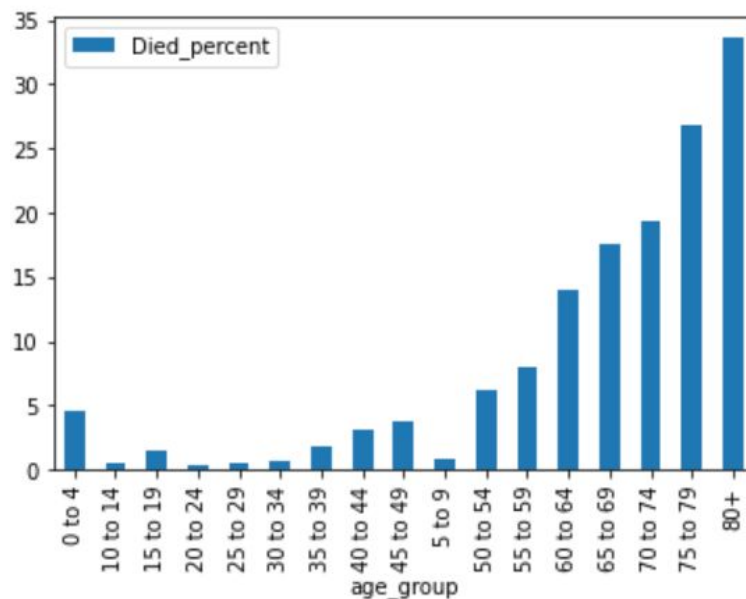
	age_group	age	status
0	35 to 39	38.0	Recovered
1	40 to 44	44.0	Died
2	60 to 64	60.0	Recovered
3	45 to 49	48.0	Recovered
4	60 to 64	62.0	Died
...
12065	35 to 39	35.0	For validation
12066	35 to 39	37.0	Died
12067	60 to 64	62.0	For validation
12068	15 to 19	18.0	Home quarantined
12069	50 to 54	50.0	For validation

12070 rows × 3 columns

Then, we extract the unique data from the status column into a new column and accumulate the value according to the corresponding age group. We also calculate the total case, died percentage and recovered percentage for each age group.

status	Admitted	Died	for_validation	home_quarantined	Recovered	total	Died_percent	Recovered_percent
age_group								
0 to 4	22	7	100	11	14	154	4.55	9.09
10 to 14	8	1	142	11	7	169	0.59	4.14
15 to 19	25	4	228	15	10	282	1.42	3.55
20 to 24	97	3	451	90	126	767	0.39	16.43
25 to 29	140	6	718	184	241	1289	0.47	18.70
30 to 34	185	11	745	230	339	1510	0.73	22.45
35 to 39	129	20	563	130	219	1061	1.89	20.64
40 to 44	128	30	506	114	189	967	3.10	19.54
45 to 49	151	38	483	103	227	1002	3.79	22.65
5 to 9	5	1	98	6	12	122	0.82	9.84
50 to 54	137	60	441	101	234	973	6.17	24.05
55 to 59	172	77	416	82	220	967	7.96	22.75
60 to 64	168	129	348	68	208	921	14.01	22.58
65 to 69	153	125	193	50	192	713	17.53	26.93
70 to 74	123	109	163	44	126	565	19.29	22.30
75 to 79	55	77	88	9	58	287	26.83	20.21
80+	72	108	90	14	37	321	33.64	11.53

The died percentage has been visualized into a graph based on the age group. Based on our analysis and from various articles, we found that older Filipinos are at the highest risk from Covid-19.



The high mortality rate of 40 percent from Covid-19 disproportionality affected the older Filipinos. As of April 15, 2020, there are 5453 confirmed covid-19 cases and 349 deaths, of which 50.7 percent are aged 65 and above. Similarly, the Malaysian health ministry established that 32 out of 99 reported of covid-19 deaths in the country are of those aged between 61 to 70, while 19 deaths involved those aged between 71 and 80 reported on May 3, 2020. In mainland China, 80 percent of all deaths occurred among people aged 60 and above as of February 11, 2020 while in the United States 8 out of 10 Covid-19 death cases have been in adults 65 years old and older.

As stated by Centers for Disease Control and Prevention (CDC) on Coronavirus Disease 2019, (July 2020), the risk of severe Covid-19 disease among adults rises with age, with the older adults at the highest risk. Severe illness means that the COVID-19 person may require treatment, intensive care, or a ventilator to help them breathe, or may even die. When you grow older, COVID-19 raises the risk for severe illness. People in their 50's, for example, are at greater risk of being severely affected.

10.5) Question 4: Which city in the Philippine has led to the increased number of cases in the Philippine?

To find which province had led to the increased number of cases in the Philippines, we used two columns to do the analysis which are case_id and province columns. Then, we have accumulated the number of provinces based on case_id.

	province	count
0	Negros Oriental	7
2	Bohol	1
3	Metropolitan Manila	6873
4	Rizal	417
12	Bulacan	141
...
7038	Ifugao	2
8321	Sulu	1
8783	Guimaras	2
9289	Davao Occidental	1

We used the max and min function to find which province has the highest Covid-19 case and minimum Covid-19 case.

```
promax = data5.max()
promax

province    Zamboanga del Sur
count                6873
dtype: object
```

```
promin = data5.min()
promin

province    Abra
count                1
dtype: object
```

Based on min and max function, we have found that Zamboanga del Sur is the province that has the highest Covid-19 case while Abra is the province that has the lowest Covid-19 case.

10.7) Case Study and K-means

With the dataset, we want to predict certain cases that could happen. We will apply the Machine Learning algorithm which is the K-Mean algorithm in order to solve these cases and assist us in our prediction.

Before we start to help the machine to predict, we process these datas. We have changed datas that are string to numeric.

For the sex column, we have categorised into 0 for male while 1 for female. For home_quarantined 0 will indicate No which means that the person did not quarantine at home, 1

means Yes, that informs us the patient did quarantine at home. For status and province we have added two new columns that are 'stat' and 'prov' that are represented as below:

	age	sex	home_quarantined	status	province	muni_city	stat	prov
0	60.0	1	0.0	Recovered	Bohol	Panglao	4	8
1	48.0	0	0.0	Recovered	Metropolitan Manila	Taguig	4	37
2	62.0	0	0.0	Died	Rizal	Cainta	1	52
3	58.0	1	0.0	Died	Rizal	Cainta	1	52
4	39.0	0	0.0	Recovered	Metropolitan Manila	Makati City	4	37
...
5227	61.0	0	1.0	For validation	Zamboanga del Sur	Zamboanga City	2	60
5228	29.0	0	1.0	Admitted	Zamboanga del Sur	Zamboanga City	0	60
5229	62.0	1	1.0	Admitted	Zamboanga del Sur	Zamboanga City	0	60
5230	45.0	0	1.0	For validation	Zamboanga del Sur	Zamboanga City	2	60
5231	18.0	1	1.0	Home quarantined	Samar	Tarangnan	3	54

Category for stat, in indicating their status:

Admitted = 0

Died = 1

For Validation = 2

Home Quarantined = 3

Recovered = 4

Category for prov, in indicating their
province::

Abra = 0

Agusan del Norte = 1

Aklan = 2

Albay = 3

Antique = 4

Bataan = 5

Batangas = 6

Benguet = 7

Bohol = 8

Bukidnon = 9

Bulacan = 10

Cagayan = 11

NAN=12

Camarines Sur = 13

Camiguin = 14

Capiz = 15

Catanduanes = 16

Cavite = 17

Cebu = 18

Compostela Valley = 19

NAN=20

Davao Oriental = 21

Davao del Norte = 22

Davao del Sur = 23

Guimaras = 24

Ifugao = 25

Ilocos Norte = 26

Ilocos Sur = 27

Iloilo = 28

Isabela = 29

La Union = 30

Laguna = 31

Lanao del Norte = 32

Lanao del Sur = 33

Leyte = 34

Maguindanao = 35

Marinduque = 36

Metropolitan Manila = 37

Misamis Occidental = 38

Misamis Oriental = 39

Negros Occidental = 40

Negros Oriental = 41

North Cotabato = 42

Northern Samar = 43

Nueva Ecija = 44

Romblon = 53

Nueva Vizcaya = 45

Samar = 54

Occidental Mindoro = 46

South Cotabato = 55

Oriental Mindoro = 47

Sultan Kudarat = 56

Palawan = 48

NAN=57

Pampanga = 49

Tarlac = 58

Pangasinan = 50

Zambales = 59

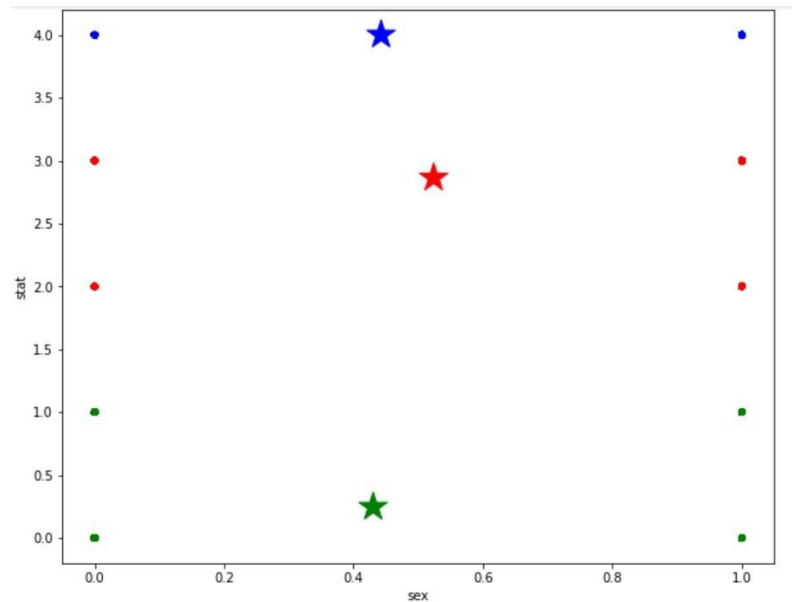
Quezon = 51

Zamboanga del Sur = 60

Rizal = 52

First case study:

There is a group that just came back to Malaysia. In the airplane, there is one person that has been diagnosed as positive Covid-19. The group has a family consisting of a couple in their 30s and their two children who are a 6-year-old girl and a baby boy, three young men in their 20s, an old man in his 60s and a woman in her 50s. From the dataset, which person has the highest possibility of being affected?



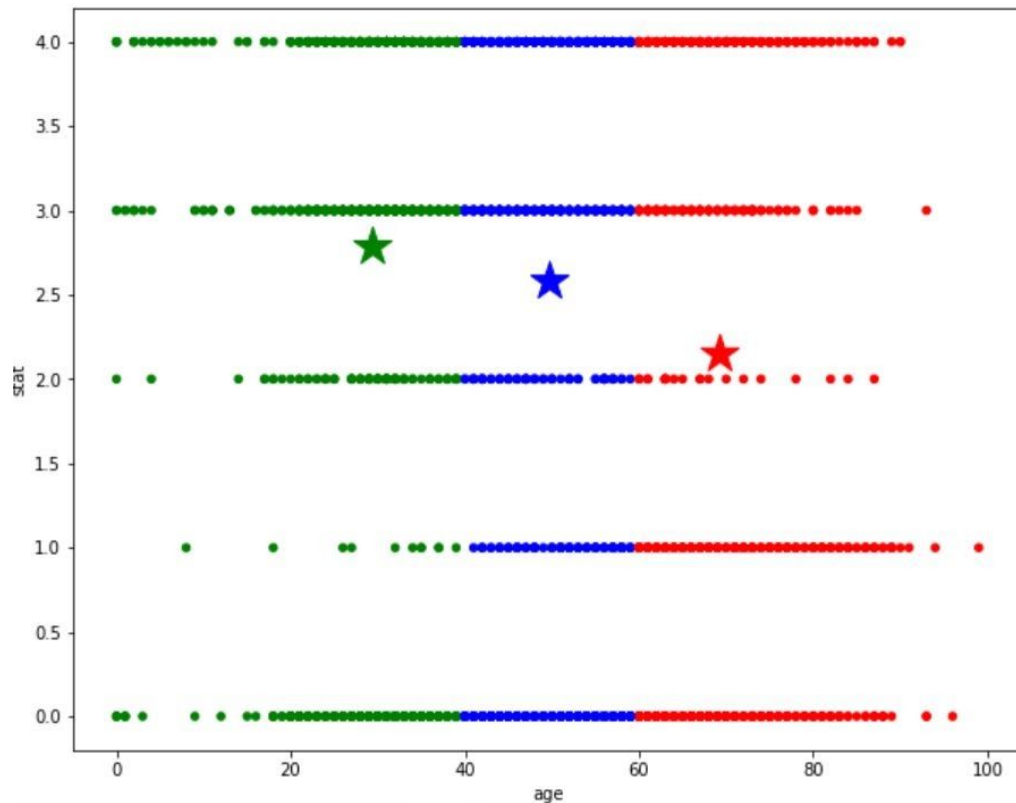
Based on this green shaped star, it categorizes status 0 and 1 in which their status are admitted and died, into two groups, one group for male and female. While for the red colored cluster, it represents status 2 and 3, which their status are for validation and home quarantined in another group and lastly the blue colored cluster status 4 means recovered.

We can not achieve the case study based on this KMeans method since the dataset is not in numeric and when we try to change, it does not work as we expected.

However based on the study about the dataset and some research on the Internet , the highest possibility of being affected with Covid-19 virus are elderly people elderly and also kids around 3 years and below. This is because their immunity systems are not strong enough to fight the virus. The other factor is also because the elderly tend to have other known medical problems that have already existed in them, whether they know it or not.

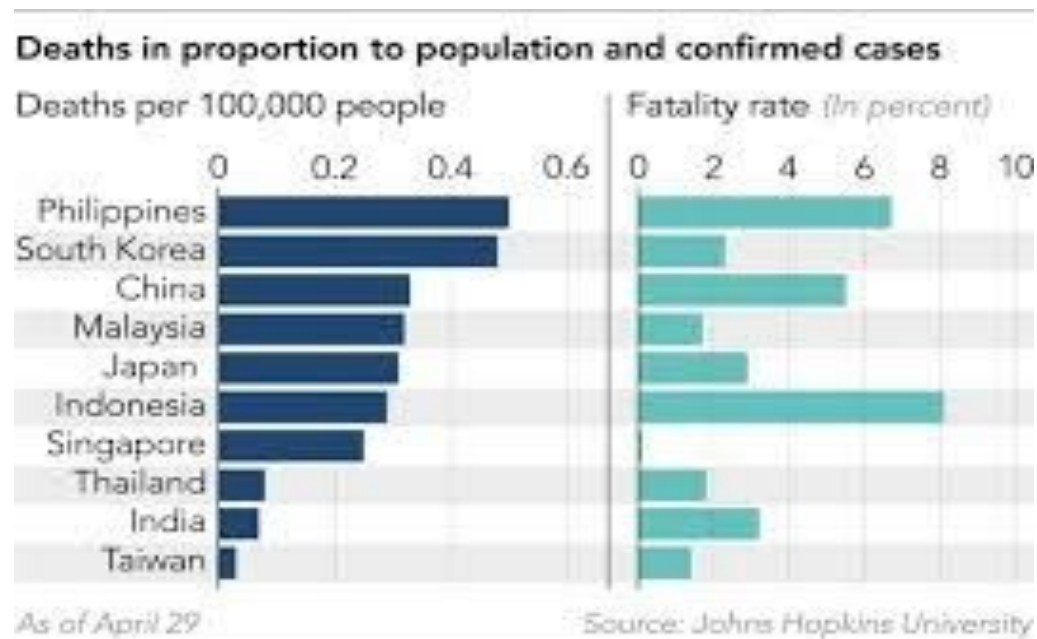
Second case study:

In the ward of a hospital that treats COVID, there was a 16-year-old girl, two 24-year-old and 27-year-old boys, an old woman aged 63, and a middle-aged man aged 46-year-old. Based on the group, who has the highest possibility of recovering from the COVID and who has the highest possibility of death?



Seen here is a graph that was used using the K-Mean algorithm, to predict the chances of recovery for the study case mentioned above. The chances for those who are aged above 40 years old are highly unrecoverable, and those aged between their 20's have a fifty percent of recovery depending on their living lifestyle, and those who are aged below that are most likely to recover from the disease. But, this does not mean that it is not risky, as certain factors like, health and their environment are also to be considered.

11.0) Comparison of Covid-19 case between Philippines and Malaysia.



In the Philippines MCO started in the metro area of Manila on March 12, a partial lockdown. But that soon was expanded to the whole Luzon region on 16th March. On 17th March the President of Philippines declared the Philippines under state of calamity for a 6 months period. On March 16th, before the state of calamity, there were 127 active cases in the Philippines. After the lockdown was eased on 15th May, active cases reached 8925. The cases kept on increasing and the new case rate also increased.

While in Malaysia, MCO was placed on May 18th when active cases were 728 and new cases rate was around 110 per day. By the end of the MCO period in Malaysia, when the country went into a recovery period on June 10th, active cases went to 1206, however, the new cases added per day were around 7. Which is a big contrast to the Philippines.

Malaysia lost fewer people to Covid-19 than the Philippines. We also have to consider things like medical facilities where Malaysia has a more robust medical sector and better public health awareness than the Philippines.

All in all, Lockdowns have vastly helped Malaysia in curbing Covid-19 cases and so did their public health policies. In the Philippines, their situation was relatively good until the movement restrictions were lifted considerably. That is what made a huge difference compared to Malaysia and they got far more cases.

12.0) References

- [illegible]

<https://www.channelnewsasia.com/news/asia/philippines-records-region-s-biggest-daily-rise-in-covid-19-12926402>

10. The Star Online. (2020, June 25). 100 days of MCO: Malaysia's success in flattening the Covid-19 infection curve. Retrieved August 1, 2020, from <https://www.thestar.com.my/news/nation/2020/06/25/100-days-of-mco-malaysia039s-success-in-flattening-the-covid-19-infection-curve>

11. Coronavirus Cases:. (n.d.). Retrieved August 01, 2020, from <http://www.worldometers.info/coronavirus>

12. Buenaventura, R., Ho, J., & Lapid, M. (2020, April 30). COVID-19 and mental health of older adults in the Philippines: A perspective from a developing country. Retrieved July 27, 2020, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7235307/>

13. Harun, H., & Othman, N. (2020, May 03). Elderly, chronically ill comprise most Covid-19 deaths in Malaysia: New Straits Times. Retrieved July 27, 2020, from <https://www.nst.com.my/news/nation/2020/05/589545/elderly-chronically-ill-comprise-most-covid-19-deaths-malaysia>

14. Older Adults and COVID-19. (n.d.). Retrieved August 02, 2020, from <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html>